Discrete Random Variables

4.1 Probability Distributions for Discrete Random Variables

This chapter will deal with the construction of discrete probability distributions by combining the methods of descriptive statistics and those of probability. Probability Distributions will describe what will probably happen instead of what actually did happen.

A **random variable** is a variable that assumes numerical values associated with the random outcomes of an experiment, where only one numerical value is assigned to each sample point.

Example: If an athlete takes two penalty kicks in a soccer match over the course of a game, she can make two, one, or no goals. The sample space is as follows:

Goal, Goal	Goal, Miss
Miss, Goal	Miss, Miss

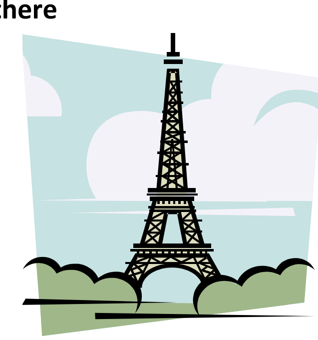
Let X = the number of goals that she makes. Then X can equal 2, 1, or 0. We can assign these values to the points in our sample space:

Goal, Goal (2)	Goal, Miss (1)
Miss, Goal (1)	Miss, Miss (0)

Now we would have a random variable x with possible values 0, 1, or 2.

There are two kinds of random variables:

- **A discrete random variable can assume a countable number of values.**
  - **Example: Number of steps walked visiting the Eiffel Tower**
- **A continuous random variable can assume any value along a given interval of a number line.**
  - **Example: The time a tourist stays at the top once s/he gets there**



Discrete random variables take on a countable number of values.

Examples of discrete random variables

- Number of sales
- Number of calls
- Shares of stock
- People in line
- Mistakes per page

<u>Continuous random variables</u> can assume any value contained in one or more intervals.

Examples of continuous random variables

- Length
- Depth
- Volume
- Time
- Weight

Example 60: Label the random variables listed below as discrete or continuous

a. The length of time customers spend waiting in line at Publix
b. The number of books purchased last year
c. The amount of weight gained by students during freshman year
d. The number of oil spills off the Alaskan coast.

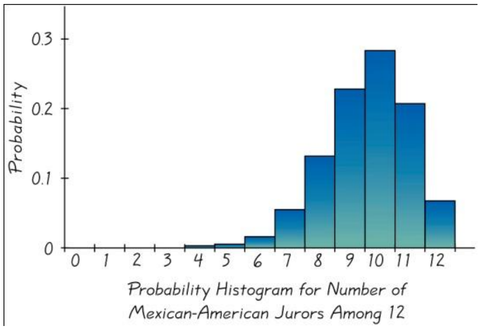
Solution: a. Continuous b. Discrete c. Continuous d. Discrete

Probability Distributions for Discrete Random Variables

This section introduces the important concept of a probability distribution, which gives the probability for each value of a variable that is determined by chance.

The **<u>probability distribution of a discrete random variable</u>** is a graph, table, or formula that specifies the probability associated with each possible value the random variable can assume.

An example of a probability histogram, one of the graphs used to represent discrete probability distributions, is given below:



Think of a probability 'distribution' as how the 100% of total probability is divided up among the possible outcomes of an experiment.

Example 61: Assume that having a boy or a girl is equally likely when having a child and derive the probability distribution for the random variable X = the number of girls when having two children.

Requirements of a probability distribution:

1. $0 \le p(x) \le 1$
2. $\sum_x p(x) = 1$

Example 62: Determine if the following is a probability distribution:

x	$P(x)$
0	0.243
1	0.167
2	0.213
3	0.149
4	0.232
5	0.164

4.2 Expected Value: The Mean of a Discrete Random Variable

The mean or **Expected Value** of a discrete random variable x is: $\mu = E(x) = \sum x \cdot p(x)$

Example 63: Below lists the probability distribution for X = the number of three-point shots made (out of 4 attempts) by Dwyane Wade during a single game. The probabilities are based on Wade's 3P percentage from the 2010 season. Find the mean of the given probability distribution.

x	$P(x)$
0	0.2320
1	0.4091
2	0.2706
3	0.0795
4	0.0088

Example 64: How much money on average will an insurance company make off of a 1-year life insurance policy worth $10,000, if they charge $290 for the policy, and you have a 0.999 probability of surviving the year?

*Note even if the company has to pay out $10,000 it keeps the $290 for itself, so a payout is only a net loss of $10,000 - $290 = $9,710.

Example 65: What is your expected value on the following game? You offer your friends $1 if they can roll a one on a die, $3 if they can roll a three and $5 if they can roll a five. However, they pay the dollar amount equal to what they roll on the die if they roll any even number on the die. Should your friends play this game against you?

Example 66: A contractor is considering a sale that promises a profit of \$38,000 with a probability of 0.7 or a loss (due to bad weather, strikes, and such) of \$16,000 with a probability of 0.3. What is the expected profit?

A. \$26,600
B. \$22,000
C. \$37,800
D. \$21,800

Solution: D. \$21,800

4.3 Standard Deviation of a Discrete Random Variable

The **population variance for a random variable** x is

$$\sigma^2 = E(x-\mu)^2 = \sum (x-\mu)^2 \cdot p(x) = \sum x^2 \cdot p(x) - \mu^2$$

The **population standard deviation for a random variable x** is given by taking the square root of the above mentioned population variance: $\sigma = \sqrt{\sigma^2}$

The rules we learned in chapter 2 can be used to describe the distribution of data on the number line within one standard deviation from the mean, two standard deviations, and so on … The table below shows the relative probabilities under the different rules:

Chebyshev's	Empirical Rule	
≥ 0	≅ .68	$P(\mu-\sigma < x < \mu+\sigma)$
≥ .75	≅ .95	$P(\mu-2\sigma < x < \mu+2\sigma)$
≥ .89	≅ .997	$P(\mu-3\sigma < x < \mu+3\sigma)$

Example 67: The following table gives the probability that x patients out of four will survive five years after receiving a diagnosis of early-stage lung cancer. Calculate the expected value for the random variable x, calculate the standard deviation, and use Chebyshev's rule to produce an interval which will capture at least 75% of the x values.

X	0	1	2	3	4
P(x)	0.049	0.220	0.372	0.280	0.079

4.4 Binomial Distribution and Binomial Probability

This section presents a basic definition of a binomial distribution along with notation, and it presents methods for finding probability values. Binomial probability distributions allow us to deal with circumstances in which the outcomes belong to two relevant categories such as acceptable/defective, success/failure, or survived/died.

Characteristics of a binomial random variable:

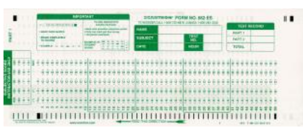
1. Experiment consists of n identical trials
   For example: Flip a coin 3 times

2. There are only two possible outcomes for each trial (success or failure)
   For example: Outcomes are Heads or Tails
3. The probability of a success remains the same from trial to trial
   For example: P(Heads) = .5; P(Tails) = 1-.5 = .5
4. The trials are independent
   For example: H on flip *i* doesn't change P(H) on flip *i* + 1
5. The random variable is the number of successes in n trials
   For example: Let x = number of heads in three flips

Example 68: Is the following a binomial experiment? A marketing firm conducts a survey to determine if consumers prefer the appearance of the bottle for Absolut Vodka over the bottle of its two top rivals. 1000 people will be asked to pick their favorite bottle, and the number of people who select the Absolut bottle will be counted. What if we instead recorded the name of the brand that was chosen by each person?

Solution: The first scenario is a binomial experiment, but the second is not.

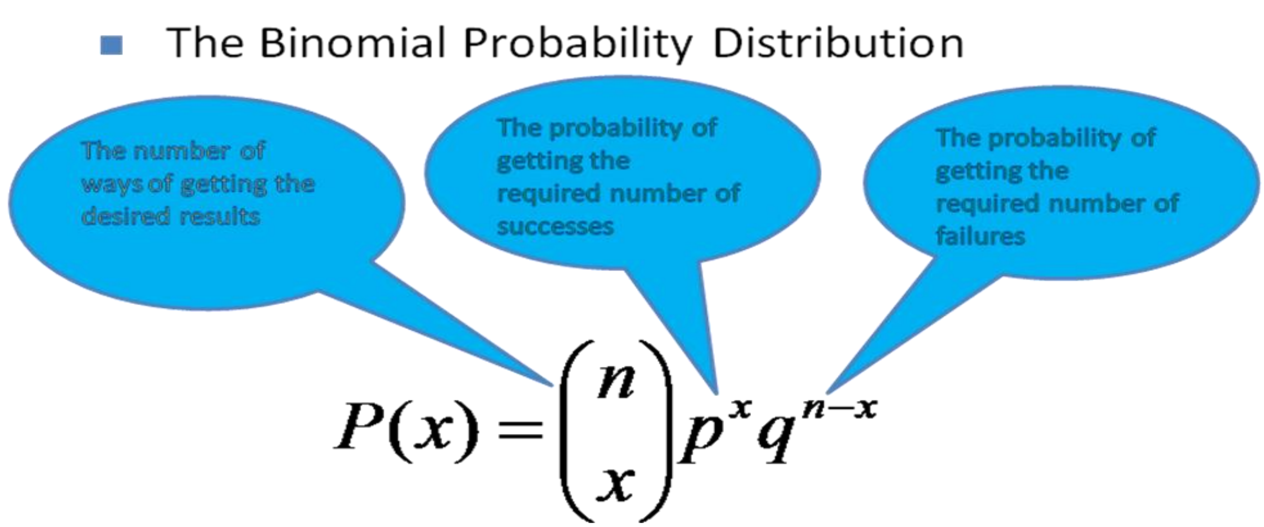
Example 69: Is the following a Binomial experiment? Let x represent the number of correct guesses on 5 multiple choice questions where each question has 5 answer options.



Solution: Yes, there are two possible outcomes correct or incorrect, a fixed number of trials (5), they are independent trials, the probability for a correct answer is 1/5 for each guess, and x counts the number of successes.

Example 70: Derive the probability distribution for the above problem, and again, let x represent the number of correct guesses on 5 multiple choice questions where each question has 5 answer options.

X	0	1	2	3	4	5
P(x)						



$$p(x) = \binom{n}{x} p^x q^{n-x}$$ for x = 0, 1, 2, . . . , n

p = probability of a success

q = 1 - p

x = number of successes

n = number of trials

Some tips when finding binomial probability:

- **Be sure that *x* and *p* both refer to the same category being called a success.**
- **When sampling without replacement, consider events to be independent if n < 0.05*N=the population sample size.***

Example 71: Say 40% of the class is female. If I randomly select ten student numbers from the roster with ***replacement***, what is the probability that exactly 6 will be female?

Solution:

$$P(x) = \binom{n}{x} p^x q^{n-x}$$
$$= \binom{10}{6} (.4^6)(.6^{10-6})$$
$$= 210(.004096)(.1296)$$
$$= .1115$$

4.5 Using the Binomial Table

Sometimes the calculations can be tedious using the binomial formula, so to speed things up, we can use a **Binomial Table.**

Example 72: Use the binomial table to confirm our calculations for the five-question multiple choice example above by finding the probability that a person gets 3 or less questions correct by guessing.

Solution: Here is a small part of a similar table:

n = 5	k	ρ = 0.15	ρ = 0.20	ρ = 0.25
	0	0.44371	0.32768	0.23730
	1	0.83521	0.73728	0.63281
	2	0.97339	0.94208	0.89648
	3	0.99777	0.99328	0.98438
This table gives the $p(x \le k)$.	4	0.99992	0.99968	0.99902
	5	1.00000	1.00000	1.00000

The answer is 0.99328.

4.6 Mean, Variance, and Standard Deviation of a Binomial Random Variable

A Binomial Random Variable has Mean, Variance, and Standard deviation:

Mean = $\mu = n \cdot p$

Variance = $\sigma^2 = n \cdot p \cdot q$

Standard Deviation = $\sigma = \sqrt{n \cdot p \cdot q}$

Example 73: A study completed in 2011 looked at the effects of self-control on relationship infidelity. The hypothesis under investigation was that people have a limited reserve of self-control, so when exercising it in one area (for example, when trying to maintain a diet), they might be less able to discipline themselves in other areas. In the study, participants currently involved in an exclusive romantic relationship and depleted of self-control by resisting cookies were more likely to make a coffee date with and disclose their personal phone number to a confederate than their non-depleted counterparts. There was no difference in self-reported likelihood of engaging in either behavior based on condition. Those in the depletion condition self-reported the same likelihood of accepting a coffee date and giving a personal phone number as those not depleted, yet when tempted in the study, the depleted group did end up cheating at a significantly higher rate than the not depleted group. The depleted group accepted an invitation to go on a date (and gave out their number) in 74% of the cases (versus only 31% for the not depleted group). Using the 74% from the study as the rate of infidelity, find the mean and standard deviation for the number of people in a committed relationship, out of a group of 49, that would commit infidelity while in a depleted state. Would it be unusual for less than 26 people in the group to be unfaithful?